DOCUMENT RESUME

ED 329 940 CS 010 492

AUTHOR Meyer, Linda A.; And Others
TITLE Assessing Early Reading with Something Old, Something New, Something Borrowed, and Something Different. Technical Report No. 529.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Apr 91
CONTRACT G0087-C1001-90
NOTE 65p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Beginning Reading; Criterion Referenced Tests; *Early Reading; Norm Referenced Tests; Primary Education; Reading Achievement; Reading Diagnosis; *Reading Tests; *Test Validity
IDENTIFIERS *Ecological Validity; Emergent Literacy; Error Detection

ABSTRACT

A study examined the ecological validity of five modified measures for children in the prereading and initial reading stages of development. Subjects, 650 students in two cohorts, kindergarten through second grade, were administered modifications of the Early Reading Test (ERT), the Chicago Reading Test (Chicago), the Error Detection Test (EDT), the Weber Test (Weber), and the Eugene Test (Eugene). Results indicated: (1) a strong relationship between the criterion-referenced Chicago test and an individually administered norm-referenced measure of letter and word identification; (2) weaker relationships, especially over time, between the ERT and tests on which children actually read words; and (3) that scores on the EDT, Weber, and Eugene tests showed that children as young as 6 years of age could detect and explain errors in what they read. (Thirteen tables of data are included; 19 references and 4 reading tests are attached.) (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED329940

CENTER FOR THE STUDY OF READING

Technical Report No. 529

ASSESSING EARLY READING WITH SOMETHING OLD, SOMETHING NEW, SOMETHING BORROWED, AND SOMETHING DIFFERENT

Linda A. Meyer
University of Illinois at Urbana-Champaign
Robert L. Linn
University of Colorado at Boulder
C. Nicholas Hastings
University of Illinois at Urbana-Champaign

April 1991

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Anderson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS010492

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90 with the Reading Research and Education Center. The publication does not necessarily reflect the views of the agency supporting the research.

BEST COPY AVAILABLE

EDITORIAL ADVISORY BOARD
1990-91

James Armstrong
Gerald Arnold
Diana Beck
Yahaya Bello
Diane Bottomley
Clark A. Chinn
Candace Clark
John Consalvi
Irene-Anna N. Diakidoy
Colleen P. Gilrane
Barbara J. Hancin
Richard Henne
Michael J. Jacobson
Carole Janisch
Bonnie M. Kerr
Paul W. Kerr
Daniel Matthews
Kathy Meyer Reimer
Montserrat Mir
Jane Montes
Juan Moran
Anne Stallman
Bryan Thalhammer
Marty Waggoner
Ian Wilkinson
Hwajin Yi

MANAGING EDITOR
Fran Lehr

MANUSCRIPT PRODUCTION ASSISTANTS
Delores Plowman
Debra Gough

Abstract

No study is better than the measures researchers use to assess its results, and the problems with the tests most commonly used by researchers for studies of children in the early elementary grades are substantial. Therefore, in the context of a longitudinal study of reading development, it was judged desirable to create new measures that were ecologically valid for children in the prereading and initial reading stages of development. This report describes the development or modification of five instruments: The Early Reading Test (ERT), The Chicago Reading Test (Chicago), The Error Detection Test (EDT), The Weber Test (Weber), and The Eugene Test (Eugene). These tests measured children's abilities to recognize (a) print from their environment (ERT); (b) letter sounds, word endings, and words (Chicago); and errors in text (EDT, Weber, and Eugene). The tests were administered to approximately 650 children in two cohorts, kindergarten through Grade 2. Results showed a strong relationship between the criterion-referenced Chicago test and an individually administered norm-referenced measure of letter and word identification. Weaker relationships were found, especially over time, between the ERT and tests on which children actually read words. Scores on the EDT, Weber, and Eugene tests showed that children as young as six could detect and explain errors in what they read. Results are discussed in reference to the relationship between fall and spring scores, the tests' characteristics in general, children's comprehension monitoring ability, and their ecological validity when measuring young children's reading abilities.

ASSESSING EARLY READING WITH SOMETHING OLD, SOMETHING NEW, SOMETHING BORROWED, AND SOMETHING DIFFERENT

Because of the mismatch between views of the reading process that now guide reading instruction and those that underlie widely used norm-referenced reading tests, there is general agreement in the field of reading that there is a profound need to develop new ways to assess student performance (Johnston, 1983; Valencia & Pearson, 1986). While educators and researchers have focused their concerns on test-related problems in the middle, upper elementary, and middle school grades, we argue that the problems are just as acute in the early elementary grades.

In fact, researchers do little testing in regular classroom settings in the early elementary grades. In kindergarten and first grade, testing usually consists of administering group measures that have children identify a word from a short series of words after having the target word read to them. It is important to keep in mind that children do not actually have to read in this test format. They simply choose the word they heard read.

There are a number of problems with this kind of testing. First, group-administered tests are not set up to capture children's individual abilities. All children are expected to complete all items. Second, it frequently takes 30 to 40 minutes to complete test administration. This is far longer than children of this age usually spend in solitary activities, even those activities that have had substantial teacher preparation time. Third, word recognition is a different task from word reading where the child produces the correct word. Children can often select one word from a list of words when they would not have been able to produce the word. Fourth, there is likely to be less variance in these tests than individually administered instruments simply because they are easier for most children.

Occasionally, by the end of first grade (or definitely by the end of second grade) assessment has changed from individual letter and word formats to short passages followed by questions. The children are expected to read these passages silently to themselves and then answer the questions. Most of these test items are carefully structured so that they lend themselves to a number of questions.

Leaders in the field of emerging literacy (Barr, 1972; Durkin, 1987; Ehri, 1987; Mason & McCormick, 1983; Sulzby, 1983) believe that if we are going to test young children, it is important to test them in areas that represent their emerging reading abilities, such as recognition of words and symbols from their natural environments, letter sounds, letter names, and oral reading of word endings and word families.

This report describes a battery of criterion-referenced measures believed to be ecologically valid for children in kindergarten, first grade, and second grade, those years when children are moving into the initial stage of reading where the emphasis is often on word recognition. It also describes how these measures relate to currently used individual and group-administered tests. The description of each measure will be followed by descriptive statistics gathered when we administered it in our longitudinal study. Finally, correlations will be presented to show the relationship between these measures and individual or group-administered norm-referenced tests we gave at the same time.

Kindergarten Measures

The Early Reading Test. Mason (1983) developed the Early Reading Test (ERT) to measure kindergarten children's abilities to recognize common words and "logos" found in their environment. It is similar to tests developed by Adams, Huggins, Starr, Rollins, Zuckerman, Nickerson, and Stevens (1980) and Ehri (1987). The ERT is designed to test stage theories of reading development postulated

by Mason (1983) and Sulzby (1983), which suggest that five-year-olds learn to recognize words and phrases, such as STOP on a stop sign, from their environments in a stage before they can recognize these same words in isolation. These researchers have also hypothesized that children will recognize these familiar words out of context, without their supporting pictures, before they will be able to read word endings, word families, or short stories made up of familiar and unfamiliar words.

The ERT has seven subtests: picture words, out-of-context words, spelling, regular word reading, make-believe word reading, story reading, and story recall. Our description of this instrument and its results will focus on a total score derived from the picture words, out-of-context words, and word reading subtests because few children scored any points on the other portions of the test.

In our study, this instrument was administered individually. The picture words subtest was administered by an examiner holding small cards with the picture words displayed (see Appendix A). Children received 2 points if they identified the word in the picture correctly, 1 point if they supplied a generic name for the item (soda pop for a Coca-Cola bottle, for example), and a 0 for other responses that were less appropriate. There were 10 items in this subtest. The second subtest required children to identify the same words that had appeared in pictures in context in the first subtest, but this time there were no context clues such as drawings of the Coke bottle. The word lists for words to read and make-believe words appear in Appendix A as well. There were 24 words in these portions of the test.

The Chicago Reading Test. Barr (1983) developed the Chicago test for use in a longitudinal study. This test has four subtests: letter sounds, word endings, word families, and a list of random words. Barr constructed this test in part because her earlier work (1972) illustrated the important relationship between children's knowledge of letter sounds and their ability to read words and sentences. The items and instructions for this instrument appear in Appendix B.

This instrument was also administered individually, with a stopping criterion. Children who failed to identify two consonants and one vowel correctly did not go on to the word reading portion of the test. It was administered to one group of children in their fourth month of kindergarten and again at the end of their kindergarten year. It was administered to a second group of children at the beginning and end of their kindergarten year as well. A total score composed simply of adding children's points for each of the subtests will be reported in this report.

Norm-referenced measures. We also administered three norm-referenced measures to the kindergarten children in our study. The Wide Range Achievement Test (WRAT) (Jastak, Jastak, & Bijou, 1978) was selected because it has been used successfully at this grade level by several researchers (Becker & Engelmann, 1977; Stevenson, Parker, Wilkinson, Hegion, & Fish, 1976), and because the reading subtest has a letter name subtest in which children identify letters by name or match uppercase letters if they have trouble identifying them. All students read aloud words in isolation until they miss 12 consecutive words. Lower performing children also identify the first two letters in their given names to complete the letter subtest. The WRAT is also administered individually. It provides a combined letter and word recognition score that can range from 1 to 100. Furthermore, the same level of the WRAT can be administered to a group of children for several consecutive years and therefore provide a repeated measure of the children's reading abilities.

The Reading Comprehension Passage subtest of the Woodcock Reading Mastery Tests (Woodcock, 1973) was selected for administration for reasons very similar to those that guided the selection of the WRAT. The Woodcock is individually administered. Children can read the phrases and passages orally. Administration stops as soon as students miss five consecutive items. Students' scores can range from 1 to 85 and thus, like the WRAT, the instrument could be used with the same group of students for several consecutive years.

The California Achievement Test (CAT) Reading subtest (CTB/McGraw-Hill, 1973) was selected for administration because it is quite representative in form and content to many group-administered kindergarten instruments. The belief was that it was important to include an instrument of this type that researchers and school personnel were familiar with in addition to individually administered instruments for purposes of comparison among measures.

First-Grade Measures

The Error Detection Test. One new measure was developed for use at the first-grade level, the Error Detection Test (EDT) (Meyer, Hastings, & Linn, 1985). It has long been accepted that six-year-olds could not monitor their own comprehension *while reading*. A review of research on comprehension monitoring (Meyer, 1986), however, revealed that this conclusion was based almost exclusively upon findings from research on *listening* comprehension. We developed the EDT to assess children's abilities to monitor their own comprehension *while reading* by identifying words that spoiled the meaning of short passages, and by telling what happened at the wrong time in sequences.

Our goal in developing the EDT was to create a criterion-referenced measure to be administered individually to first and second graders. All reading vocabulary in this instrument had appeared in the reading or science textbooks used in the school districts of the children tested. Therefore, the children had a reasonably good chance of having been exposed to the words in the items. To further the children's opportunities to perform well on the comprehension aspects of the test, they read aloud and were corrected when they made oral reading errors.

The EDT has two subtests. The first is called Absurd Target Words, the second, Impossible Sequence. On the first subtest, examiners correct students' word reading errors by telling them the correct word and coding each misidentification, deletion, insertion, and self-correction. After a student reads a passage, such as "Leaves turn red. Leaves turn yellow. Leaves turn blue," she is asked to tell which word spoiled the meaning. If the child correctly identifies the absurd target word, she is then asked, "Why?"

The second subtest again requires students to read aloud then to tell "What happened at the wrong time." If the student correctly identifies what is out of sequence, he is asked to explain why. Thus, there are three scores for each of these three subtests, (a) the decoding score, which is the number of words missed in all of the items comprising a subtest; (b) a score for identification of the absurd target words or the out-of-sequence events; and (c) support for the absurd target word or sequence selected. A copy of the EDT appears in Appendix C.

Norm-referenced measures. The WRAT decoding subtest and the Reading Comprehension Passage subtest of the Woodcock were administered, with the EDT, to first-grade students in the spring.

Second-Grade Measures

The EDT was administered again to all students in the fall of their second-grade year. The WRAT and the Woodcock were also administered at the beginning of second grade to both groups of children. In addition, two new measures were added to the second-grade battery, the Weber Test and the Eugene Test.

The Weber Test. This measure of comprehension monitoring was developed by Weber (1971) to assess inner-city students' comprehension abilities. It is group-administered and takes about 20 minutes. The test was originally designed for use with third-grade students. We were granted permission to use the instrument with the stipulation that we would not circulate copies of it. Therefore, the following description of the measure is only suggestive of Weber's original test.

Each passage is approximately 15-20 words long. These words usually make up two to three sentences. The content of the item "builds up" so that the insertion of an absurd word at or close to the end is fairly obvious. There is a practice item, and the children are told to circle the word that "spoils the meaning." Items follow this format: "George's father is a baker. He bakes cakes and pies and other things. He helps the men who play with him." There were 32 items in the original third-grade instrument that we divided into two instruments for piloting. We then selected items from each form to yield a new 22-item form.

The Eugene Test. This instrument was originally developed by Engelmann and Meyer (1973) as a series of items for second- and third-grade students to use for practice in preparation for standardized test taking. It is administered to whole classes of students and takes approximately 30 minutes. The first portion of the test is a set of riddles. The other items have made-up words in them. The intent of this practice was to help students learn that they could figure out answers on tests even if they were not familiar with all of the words in the text. Thus, we believed that these items were also good tests of children's comprehension monitoring ability. A copy of the complete instrument appears in Appendix D.

We next present the descriptive and correlational results from the administration of these instruments to approximately 650 children in two cohorts one year apart in school. The children were between five and six years of age when they began school. They were followed longitudinally from kindergarten through second grade.

Descriptive Results

Kindergarten

Table 1 lists the means, standard deviations, alphas, and skewness of the kindergarten measures administered to the students. Cohort 1 designates the first group of students to take these tests, Cohort 2, the second group. The "S" indicates measures administered in the spring, "F," tests that were given in the fall. The ERT was administered only to Cohort 1. The Chicago was administered first to Cohort 1 in the winter, whereas it was administered at the beginning of the school year to Cohort 2. With these exceptions, the measures were administered at comparable times to both cohorts.

On average, the entering kindergartners had been characterized as being able to identify a few environmental print words or their generic counterparts, most letters by letter name, and a few letter sounds. By spring, these same students could identify many letter sounds and some word endings, all letters on the WRAT, and they could read a few words. They could also figure out a few short cloze comprehension passage items on the Woodcock and correctly select almost all of the 21 words read to them on the CAT. In addition, Cohort 2 performed slightly higher than Cohort 1 on each instrument administered at comparable times.

[Insert Table 1 about here.]

First Grade

Table 2 presents the means, standard deviations, alphas, and skewness of the new measures, the WRAT, and the Woodcock administered to first graders. The WRAT scores appear for fall and spring, the Chicago for fall only, and the EDT and Woodcock scores for spring only.

There are six EDT scores in this table. ED DEC ATW is the number of words students decoded incorrectly while reading the Absurd Target Word items. Students read aloud 107 words in this subtest. ED ATW ID is the number of target words the student identified correctly. The highest possible score

here is 10. ED ATWS represents students' ability to support the word they identified as having spoiled the meaning of the passage. Students in both cohorts could support almost all of their choices.

ED DEC IS designates the number of decoding errors students made while reading the Impossible Sequence items. There were 116 words in these six items. Students averaged fewer correct responses on these items than on the Absurd Target Word items. In this part of the test they were asked, "What happened at the wrong time?" Both cohorts averaged almost two-thirds of the items correct. Cohort 1 supported (ED ISS) more of their choices than did Cohort 2.

On the WRAT, Chicago, and Woodcock, Cohort 2 continued to perform slightly higher than Cohort 1, just as they had done in kindergarten. On the curriculum-specific instrument, the EDT, however, Cohort 1's performance was overall superior to the performance of Cohort 2 since they missed fewer words in both sets of passages, identified more of the absurd target words, and supported their choices better.

[Insert Table 2 about here.]

Second Grade

Table 3 lists the means, standard deviations, alphas, and skewness of the new measures administered in second grade. The EDT instrument was administered in the fall. In addition, the Eugene was administered to both cohorts and the Weber to Cohort 1 in the spring. Cohort 2 made fewer decoding errors than Cohort 1 on both subtests of the Error Detection Test. Cohort 1 identified and supported more of the absurd target words, and Cohort 2 performed slightly better than Cohort 1 when identifying and supporting the impossible sequences.

[Insert Table 3 about here.]

The Eugene has 39 total possible points and the Weber 22 possible points. Table 3 shows Cohort 1 superior in performance to Cohort 2 students on the Eugene. The high scores on the Weber with Cohort 1 suggested potential ceiling effects. For this reason, the Weber was not administered to Cohort 2.

Correlational Results

Kindergarten

The next six tables show Pearson product moment correlations for each set of instruments administered to each cohort. Table 4 presents correlations for all tests given to Cohort 1. Administrations of the WRAT, Chicago, and Woodcock show high positive correlations with each other. The lowest correlations between instruments are for the ERT and the CAT, and even these are statistically significant.

[Insert Table 4 about here.]

Table 5 shows the correlations for Cohort 2 on kindergarten measures. Generally, these correlational patterns replicate those found for Cohort 1 with high correlations between the Chicago, WRAT, and Woodcock despite the different administration times for the Chicago. These tests correlated moderately with the CAT. Of particular interest are the correlations between the spring end-of-kindergarten decoding and comprehension measures. The WRAT and the Woodcock correlate .70 and the Chicago and the Woodcock correlate .64 for Cohort 1. These same instruments correlated .79 and .55 for Cohort

2. These results suggest a strong relationship between children's word recognition and comprehension abilities at the end of kindergarten.

[Insert Table 5 about here.]

First Grade

Table 6 shows the correlations found for Cohort 1 first-grade measures. The fall Chicago and WRAT scores correlate very highly (r = .76), and the EDT decoding scores also correlate moderately with the other decoding measures. The negative correlation for the EDT decoding score is the result of the way it was scored. On this test, we counted the number of *errors* students made on the two subtests, while the decoding scores on the WRAT and Chicago represent the number of words students got correct. Identification and support for both absurd target words and impossible sequences correlate statistically significantly, though the correlations are low with fall decoding measures, and at least moderately with the other spring measures.

[Insert Table 6 about here.]

Table 7 is the Cohort 2 first-grade correlation table. Overall, Table 7 replicates the findings in Table 6 for Cohort 1 first graders. While the patterns are generally the same, the correlations are moderate for the EDT identification and support subtest scores for Cohort 2 whereas they were low but significant for Cohort 1. For both cohorts the correlations for spring WRAT and Woodcock scores are very high (r = .84 for Cohort 1 and r = .79 for Cohort 2).

[Insert Table 7 about here.]

Spring of First Grade and Fall of Second Grade

Tables 8 and 9 show the spring to fall correlations for one measure, the EDT for Cohort 1 and Cohort 2. These results are quite similar for both cohorts. Correlations for decoding are comparable for both cohorts (r = .76 and r = .74) whereas Absurd Target Word identification and support are slightly more highly correlated than Impossible Sequence identification and support. These results, from an instrument administered in the spring of first grade and again in the fall of second grade, give a measure of stability of student performance for these groups of students.

[Insert Tables 8 and 9 about here.]

Second Grade

Table 10 presents the correlations for the six EDT subtests administered in the fall of second grade with the two comprehension monitoring instruments, the Eugene and the Weber, administered in the spring. Once again, all of these relationships are significant and produced moderate to high correlations. The Weber and the Eugene are highly correlated (r = .68) as one would hope, for two measures of comprehension monitoring administered at the same time during the second-grade year. The decoding subtest scores on the EDT remain very highly correlated (r = .97), as one would anticipate for instruments developed from reading and science vocabulary taught to all students. The Eugene and Weber scores have almost identical correlations with decoding the absurd target words (r = -.58 and r = -.60) and with decoding the impossible sequences (r = -.61 and r = -.62) of the EDT. These scores also show a high relationship between children's abilities to identify correctly words they have been taught and to detect errors in passages they read at the end of second grade.

[Insert Table 10 about here.]

Table 11 is identical to Table 10 except that it lacks the Weber since it was not administered to Cohort 2 at the end of the second-grade year. Once again, Cohort 2 results generally replicate Cohort 1 results on these instruments except that Cohort 2 correlations for identification and support of both Absurd Target Words and Impossible Sequences are overall lower (though still significant) for Cohort 2 when compared to Cohort 1 results. The correlations for the Eugene and the EDT subtests are close, though often lower, than the correlations for Cohort 1.

[Insert Table 11 about here.]

Correlations of Selected Kindergarten, First- and Second-Grade Measures

Tables 12 and 13 show correlations of several of the kindergarten, first-, and second-grade measures chosen to illustrate the variety of instruments administered during these three grades. In Table 12, all but two of the correlations are significant at the .001 level. This table shows generally high correlations for entering kindergarten WRAT scores and the spring second-grade scores on the Eugene (r = .47). The spring first grade scores on the WRAT (r = .59) and the Woodcock (r = .60) are also highly correlated with the end of second-grade comprehension monitoring instrument, the Eugene. Correlations between the ERT and the EDT, the Weber, and Eugene are lower than the corresponding correlations of entering scores on the WRAT or the Chicago spring scores with the EDT, Weber, and Eugene, yet they are significant. Furthermore, the first-grade Woodcock scores are highly correlated with the Weber (r = .52) and the Eugene (r = .60).

[Insert Table 12 about here.]

Table 13 again replicates the findings for the same measures with Cohort 1. The correlations tend to be a few points higher for Cohort 2 than they were for Cohort 1.

[Insert Table 13 about here.]

Discussion

This discussion addresses five questions: (a) What is the relationship between spring and fall scores on the same measures? (b) Do the measures provide evidence that young children can monitor their reading comprehension? (c) How might we explain the lower correlations between the ERT and the CAT as compared to other measures? (d) What is the relationship between kindergarten and second-grade decoding and comprehension measures? and (e) What are the contributions from these new measures?

What is the relationship between spring and fall scores on the same measures?

The kindergarten and first-grade scores on the WRAT and Chicago rose consistently from spring to fall for both Cohort 1 and Cohort 2. In a similar vein, generally, there were drops in decoding errors during that same time period for both Absurd Target Words and Impossible Sequences on the EDT. In addition, there were consistent rises in the scores for both identification and support for the Absurd Target Words and Impossible Sequences for both cohorts, spring to fall.

These findings suggest that once students reach a certain level of reading proficiency, they continue to improve their performance during the summer even in the absence of school-based instruction. These findings also suggest that the time-honored practice of dropping back in curricula to review the previous year's work before moving ahead again might be re-examined. It appears from these results in the early elementary grades that children should be moved forward as soon as school begins, instead of spending time going over work that is too easy for them.

Do the measures provide evidence that young children can monitor their reading comprehension?

Simply put, the answers to this question is yes, there is evidence that children can monitor their reading comprehension. Three instruments described in this report, the EDT, the Weber, and the Eugene, were designed to assess comprehension monitoring. The more than 600 children in our two cohorts have demonstrated that even as first graders, they could readily perform error detection tasks while reading aloud.

This finding is contrary to the lore on children's comprehension monitoring that suggests young children cannot monitor their own listening comprehension much less their own reading comprehension and that they must be older before it is even appropriate to assess comprehension monitoring. Our results suggest consistently that children are capable of monitoring their reading in the very early grades and that they develop this ability as they learn to recognize words and understand their meanings.

How might we explain the lower correlations between the ERT and the CAT as compared to other measures?

There may be several explanations for these results. First, the ERT was scored in such a way that students received partial credit for identifying the "generic" or "logo" in the Picture Word and Word Score portions of the instruments. In other words, if students said "soda pop" when shown a picture of a Coca-Cola bottle, they received 1 point. In all of the other measures administered, only exact responses were scored as correct. The ERT attempts to measure different skills than those measured by the other tests. It is assumed that children have learned the words and logos on the ERT from naturalistic exposure during the first few years of their lives. The other instruments have been developed by selecting vocabulary common in the children's school environments.

The CAT is more similar to the ERT than it is to other measures because students did not have to read words correctly in order to identify them correctly. They had only to be able to pick out the first sound of the word when they had a list of three letters read to them. In all cases, the correct choices were fairly clear if the children could identify the picture correctly. Children chose between an *n*, *c*, and *s* for the picture of a cat, and an *a*, *k*, *w*, and *e* for the picture of an eagle, for example. The high scores for most children on these 21 items in comparison to their relatively low word reading scores on the WRAT is another indication of the discrepancy in performance for reading production tasks versus performance on word recognition tasks. In short, the ERT and CAT test different things than do the other measures, and they appear to be easier.

What is the relationship between kindergarten and second-grade decoding and comprehension measures?

The high correlations between kindergarten decoding and comprehension instruments were also found at the first-grade and second-grade levels. The relationships remained strong when correlating kindergarten to second-grade instruments as well. These correlations certainly suggest a strong link between decoding and comprehension ability in the early elementary grades. Children who identify words correctly are very likely to understand what those words mean. The relationship between decoding ability and comprehension ability is powerful and lasting. It endures during at least three years of schooling.

What are the contributions from these new measures?

Taken together, these five measures show promise. They provide a sequence of tests to measure early reading. They also give teachers and researchers alternatives for measuring children's emerging literacy in ways that consider both naturalistic development and classroom instruction. The ERT appears to be sensitive to words and logos children have absorbed from their environments before they began school.

The Chicago reflects the emphasis on letter sounds and word endings found in many classrooms. It has proven to be a particularly sensitive instrument. The Error Detection, Weber, and Eugene Tests offer evidence that children can identify errors in texts thereby presenting proof of comprehension monitoring. This battery yields evidence of children's emerging literacy that could be helpful for instruction and evaluation.

They also show promise because by second grade, there are two group-administered instruments that have good reliability and validity. We recognize the importance of developing instruments that others can administer in reasonable amounts of time to groups of children once they are in second or third grade. Future research should concentrate on the development of instruments that represent new paradigms in reading as well as administrative manageability.

References

Adams, M., Huggins, A., Starr, B., Rollins, A., Zuckerman, L., Nickerson, R., & Stevens, K. (1980). *A prototype test of decoding skills* (Report No. 4316). Cambridge, MA: Bolt, Beranek & Newman.

Barr, R. (1972). The influence of instructional conditions on word recognition errors. *Reading Research Quarterly, 7*, 509-529.

Barr, R. (1983). *Chicago reading test*. Unpublished test, National College of Education, Evanston, IL.

Becker, W., & Engelmann, S. (1977). *The Oregon direct instruction model*. Eugene: University of Oregon Follow Through Project.

CTB/McGraw Hill (1973). *The California achievement test*. Monterey, CA: Del Monte Research Park.

Durkin, D. (1987). Testing in the kindergarten. *Reading Teacher, 40*, 766-770.

Ehri, L. (1987). Learning to read and spell words. *Journal of Reading Behavior, 19*, 5-31.

Engelmann, S. E., & Meyer, L. A. (1973). *The Eugene*. Unpublished test, University of Illinois, Center for the Study of Reading, Urbana-Champaign.

Jastak, J. F., Jastak, S., & Bijou, S. (1978). *The wide range achievement test*. Wilmington, DE: Jastak Associates.

Johnston, P. (1983). *Reading comprehension assessment*. Newark, DE: International Reading Association.

Mason, J. M. (1983). *Early reading test*. Unpublished test, University of Illinois, Center for the Study of Reading, Urbana-Champaign.

Mason, J. M., & McCormick, C. (1983, April). *When and how might young children be instructed about reading*? Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Meyer, L. A. (1986). *A review of research on comprehension monitoring parsed into Chall's six stages of reading development.* Unpublished manuscript, University of Illinois, Center for the Study of Reading, Urbana-Champaign.

Meyer, L. A., Hastings, C. N., & Linn, R. L. (1985). *The error detection test*. Unpublished test, University of Illinois, Center for the Study of Reading, Urbana-Champaign.

Stevenson, H., Parker, T., Wilkinson, A., Hegion, A., & Fish, E. (1976). Longitudinal study of individual differences in cognitive development and scholastic achievement. *Journal of Educational Psychology, 68*, 377-400.

Sulzby, E. (1983). *Beginning readers' developing knowledges about written language* (Final Report to the National Institute of Education, NIE-G-80-0176). Evanston, IL: Northwestern University.

Valencia, S., & Pearson, P. (1986). *New models for reading assessment* (Reading Ed. Rep. No. 71). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Weber, G. (1971). Unpublished test of error detection. In *Inner-city children can be taught to read: Four successful schools* (Occasional Papers No. 18). Washington, DC: Council for Basic Education.

Woodcock, R. W. (1973). *Woodcock reading mastery tests*. Circle Pines, MN: American Guidance Service.

Table 1

Descriptive Statistics for Kindergarten Measures
Cohorts 1 & 2

	Cohort 1					Cohort 2				
Measure	*N*	$\bar{X}$	*SD*	Alpha	Skewness	*N*	$\bar{X}$	*SD*	Alpha	Skewness
ERT, F	330	24.52	(9.68)	.47	0.48	—	—	—	—	—[a]
WRAT, F	330	18.80	(7.43)	.70	-0.38	304	20.09	(8.76)	.75	1.06
CHICAGO, F	—	—	—	—	—[a]	304	5.99	(12.68)	.72	3.04
CHICAGO, W	316	17.79	(18.20)	.73	1.06	—	—	—	—	—[a]
CHICAGO, S	331	30.74	(25.28)	.84	.69	294	32.03	(24.81)	.83	.71
WRAT, S	331	27.28	(6.79)	.75	.70	294	27.33	(8.78)	.70	1.67
WOODCOCK, S	331	2.33	(3.59)	—	2.51	294	3.18	(6.22)	—	4.18
CAT, S	331	17.57	(4.26)	.89	-1.44	295	17.96	(4.06)	.89	-1.71

[a]Instrument not administered to this cohort of students.

Table 2

Descriptive Statistics for First-Grade Measures
Cohort 1 & 2

	Cohort 1					Cohort 2				
Measure	*N*	$\bar{X}$	*SD*	Alpha	Skewness	*N*	$\bar{X}$	*SD*	Alpha	Skewness
WRAT, F	322	28.64	(7.74)	.73	2.01	300	29.25	(8.21)	.74	1.44
CHICAGO, F	322	32.05	(25.01)	.85	.65	300	35.99	(25.45)	.87	0.61
ED DEC ATW, S	320	14.92	(19.96)	.96	3.34	283	15.20	(22.69)	.97	2.80
ED ATW ID, S	320	8.83	(1.58)	.74	-3.24	283	8.80	(1.08)	.49	-2.73
ED ATWS, S	320	8.34	(2.04)	.78	-2.25	283	8.01	(1.88)	.71	-1.74
ED DEC IS, S	318	19.40	(22.47)	.96	2.98	283	20.04	(24.82)	.96	2.56
ED IS, ID, S	318	3.73	(1.46)	.48	-0.51	283	4.14	(1.35)	.42	-0.89
ED ISS, S	318	3.06	(1.53)	.55	-0.03	283	2.77	(1.45)	.53	0.02
WRAT, S	320	48.83	(9.46)	.67	0.02	283	49.72	(10.21)	.67	-0.18
WOODCOCK, S	321	21.83	(11.40)	—	0.66	283	22.82	(13.12)	.94	0.90

Table 3

Descriptive Statistics for Second-Grade Measures
Cohorts 1 & 2

	Cohort 1					Cohort 2				
Measure	N	$\bar{X}$	SD	Alpha	Skewness	N	$\bar{X}$	SD	Alpha	Skewness
ED DEC ATW, F	320	14.14	(20.17)	.96	3.14	284	12.38	(19.94)	.96	3.37
ED ATW ID, F	320	8.96	(1.27)	.66	-4.03	284	8.86	(1.19)	.59	-2.99
ED ATWS, F	320	8.56	(1.64)	.69	-2.69	284	8.40	(1.57)	.66	-2.14
ED DEC IS, F	320	20.40	(23.04)	.96	2.58	284	18.07	(22.05)	.96	2.91
EDIS ID, F	320	4.05	(1.44)	.52	-0.53	284	4.37	(1.34)	.48	-0.87
EDIS S, F	320	3.08	(1.57)	.63	0.09	284	3.11	(1.54)	.58	-0.07
EUGENE	306	27.72	(8.32)	.92	-0.60	276	25.26	(9.29)	.93	-0.30
WEBER	306	18.20	(5.52)	.95	-1.82	—	—	—	—	—[a]

[a]Instrument not administered to this cohort of students.

Table 4

Correlations of Kindergarten Measures[a]
Cohort 1

	EARLY READING TEST	WRAT, F	CHICAGO, W	WRAT, S	CHICAGO, S	WOODCOCK, S	CAT, S
EARLY READING TEST	1.00						
WRAT, F	.55	1.00					
CHICAGO, W	.40	.62	1.00				
WRAT, S	.49	.68	.71	1.00			
CHICAGO, S	.43	.61	.82	.78	1.00		
WOODCOCK, S	.38	.51	.58	.70	.64	1.00	
CAT, S	.40	.57	.51	.61	.63	.31	1.00

[a]Lowest N = 304. Correlations above .19 are significant at the < .001 level.

Table 5

Correlations of Kindergarten Measures[a]
Cohort 2

	CHICAGO, F	WRAT, F	CHICAGO, S	WRAT, S	WOODCOCK, S	CAT, S
CHICAGO, F	1.00					
WRAT, F	.67	1.00				
CHICAGO, S	.52	.64	1.00			
WRAT, S	.65	.80	.72	1.00		
WOODCOCK, S	.65	.64	.55	.79	1.00	
CAT, S	.28	.63	.60	.56	.25	1.00

[a]Lowest N = 272. Correlations above .20 are significant at the < .001 level.

Table 6

Correlations of First-Grade Measures[a]
Cohort 1

	WRAT, F	CHICAGO, F	ED DECATW, S	ED ATW ID, S	ED ATWS, S	ED DECIS, S	ED IS ID, S	ED ISS, S	WRAT, S	WOODCOCK, S
WRAT, F	1.00									
CHICAGO, F	.76	1.00								
ED DECATW, S	-.49	-.47	1.00							
ED ATW ID, S	.25	.25	-.50	1.00						
ED ATWS, S	.27	.31	-.51	.83	1.00					
ED DECIS, S	-.49	-.49	.92	-.51	-.50	1.00				
ED IS ID, S	.23	.26	-.28	.43	.43	-.29	1.00			
ED ISS, S	.22	.25	-.31	.43	.50	-.29	.80	1.00		
WRAT, S	.66	.65	-.66	.35	.38	-.68	.30	.31	1.00	
WOODCOCK, S	.69	.69	-.60	.34	.38	-.60	.31	.31	.86	1.00

[a]Lowest N = 294. Correlations above .19 are significant at the < .001 level.

Table 7

Correlations of First-Grade Measures[a]
Cohort 2

	WRAT, F	CHICAGO, F	ED DECATW, S	ED ATW ID, S	ED ATWS, S	ED DECIS, S	ED IS ID, S	ED ISS, S	WRAT, S	WOODCOCK, S
WRAT, F	1.00									
CHICAGO, F	.78	1.00								
ED DECATW, S	-.53	-.51	1.00							
ED ATW ID, S	.38	.33	-.47	1.00						
ED ATWS, S	.34	.33	-.49	.72	1.00					
ED DECIS, S	-.54	-.53	.98	-.48	-.49	1.00				
ED IS ID, S	.29	.31	-.40	.45	.48	-.39	1.00			
ED ISS, S	.30	.35	-.40	.37	.46	-.41	.68	1.00		
WRAT, S	.67	.65	-.75	.41	.49	-.78	.37	.45	1.00	
WOODCOCK, S	.64	.70	-.62	.39	.47	-.65	.35	.45	.79	1.00

[a]Lowest N = 269. Correlations above .20 are significant at the < .001 level.

Table 8

Correlations of Spring of First-Grade and Fall of Second-Grade Measures Cohort 1[a]

	ED DEC ATW, F2	ED ATW ID, F2	ED ATWS, F2	ED DEC IS, F2	ED IS ID, F2	FDS, F2
ED DEC ATW, S1	.76	-.35	-.41	.76	-.23	-.22
ED ATW ID, S1	-.40	.33	.33	-.36	.12	.12
ED ATWS, S1	-.43	.33	.37	-.39	.13	.11
ED DEC IS, S1	.76	-.29	-.36	.78	-.24	-.23
ED IS ID, S1	-.32	.16	.15	.32	.24	.27
ED S, S1	-.29	.21	.22	-.29	.24	.24

[a]Lowest N = 285. Correlations above .20 are significant at the < .001 level.

Table 9

Correlations of Spring of First-Grade and Fall of Second-Grade Measures Cohort 2[a]

	ED DEC ATW, F2	ED ATW ID, F2	ED ATWS, F2	ED DEC IS, F2	ED IS ID, F2	EDS, F2
ED DEC ATW, S1	.74	-.24	-.39	.72	-.19	-.39
ED ATW ID, S1	-.54	.29	.40	-.52	.20	.32
ED ATWS, S1	-.53	.30	.47	-.52	.27	.46
ED DEC IS, S1	.74	-.25	-.40	.72	-.19	-.38
ED IS ID, S1	-.38	.15	.26	-.36	.17	.34
ED S, S1	-.39	.23	.32	-.39	.27	.51

[a]Lowest N = 235. Correlations above .22 are significant at the < .001 level.

Table 10

Correlations of Second-Grade Measures[a]
Cohort 1

	WRAT,F	ED DEC ATW, F	ED ATW ID, F	ED ATWS, F	ED DEC IS, F	EDIS ID, F	EDIS S, F	EUGENE, S	WEBER, S
WRAT, F	1.00								
ED DEC ATW, F	-.72	1.00							
ED ATW ID, F	.34	-.49	1.00						
ED ATWS, F	.39	-.54	.86	1.00					
ED DEC IS, F	-.76	.97	-.47	-.51	1.00				
EDIS ID, F	.37	-.41	.35	.41	-.42	1.00			
EDIS S, F	.35	-.36	.32	.40	-.38	.75	1.00		
EUGENE, S	.62	-.58	.36	.42	-.61	.33	.43	1.00	
WEBER, S	.58	-.60	.39	.41	-.62	.32	.30	.68	1.00

[a]Lowest N = 288. Correlations above .19 are significant at the < .001 level.

Table 11

Correlations of Second-Grade Measures[a]
Cohort 2

	WRAT, F	ED DEC ATW, F	ED ATW ID, F	ED ATWS, F	ED DEC IS, F	EDIS ID, F	EDIS S, F	EUGENE, S
WRAT, F	1.00							
ED DEC ATW, F	-.72	1.00						
ED ATW ID, F	.32	-.30	1.00					
ED ATWS, F	.42	-.41	.81	1.00				
ED DEC IS, F	-.74	.97	-.31	-.42	1.00			
EDIS ID, F	.34	-.28	.28	.34	-.29	1.00		
EDIS S, F	.46	-.39	.32	.45	-.40	.69	1.00	
EUGENE, S	.61	-.52	.32	.46	-.55	.34	.51	1.00

[a]Lowest N = 262. Correlations above .20 are significant at the < .001 level.

Table 12

Correlations of Selected Kindergarten, First-, and Second-Grade Measures[a]
Cohort 1

	WRAT, FK	ERT, FK	CHICAGO, SK	WRAT, S1	WOODCOCK, S1	WRAT, FS	ED DEC ATW, FS	ED ATW ID, FS	ED ATWS, FS	ED DECIS, FS	EDIS, IDFS	EDIS S, FS	WEBER	EUGENE
WRAT, FK	1.00													
ERT, FK	.55	1.00												
CHICAGO, SK	.61	.43	1.00											
WRAT, S1	.51	.30	.68	1.00										
WOODCOCK, S1	.54	.35	.71	.74	1.00									
WRAT, FS	.51	.36	.64	.87	.82	1.00								
ED DEC ATW, FS	-.44	-.33	-.46	-.66	-.59	-.72	1.00							
ED ATW ID, FS	.26	.21	.18	.25	.20	.34	-.48	1.00						
ED ATWS, FS	.36	.30	.29	.35	.30	.39	-.54	.84	1.00					
ED DEC IS, FS	-.46	-.34	-.50	-.70	-.63	-.76	.97	-.46	-.51	1.00				
EDIS, IDFS	.28	.22	.24	.32	.33	.37	-.41	.33	.39	-.41	1.00			
EDIS S, FS	.37	.28	.26	.34	.36	.35	-.37	.32	.40	-.38	.75	1.00		
WEBER, SS	.39	.21	.37	.54	.52	.62	-.60	.38	.42	-.62	.31	.30	1.00	
EUGENE, SS	.47	.31	.47	.59	.60	.58	-.58	.36	.42	-.61	.32	.43	.68	1.00

[a]Lowest N = 264. All correlations above .20 are significant at the < .001 level.

Table 13

Correlations of Selected Kindergarten, First-, and Second-Grade Measures[a]
Cohort 2

	WRAT, FK	CHICAGO, SK	WRAT, S1	WOODCOCK, S1	WRAT, FS	ED DEC ATW, FS	ED ATW ID, FS	ED ATWS, FS	ED DEC IS, FS	ED IS IDFS	EDIS S, FS	EUGENE
WRAT, FK	1.00											
CHICAGO, SK	.65	1.00										
WRAT, S1	.63	.66	1.00									
WOODCOCK, S1	.60	.67	.79	1.00								
WRAT, FS	.66	.59	.91	.81	1.00							
ED DEC ATW, FS	-.49	-.39	-.70	-.59	-.72	1.00						
ED ATW ID, FS	.32	.26	.27	.29	.32	-.30	1.00					
ED ATWS, FS	.39	.28	.36	.39	.42	-.41	.81	1.00				
ED DEC IS, FS	-.52	-.41	-.71	-.62	-.74	.--	-.32	-.42	1.00			
ED IS IDFS	.31	.28	.30	.36	.34	-.29	.28	.34	.--	1.00		
EDIS S, FS	.38	.40	.44	.53	.46	-.39	.32	.46	-.41	.69	1.00	
EUGENE	.52	.41	.57	.66	.61	-.52	.32	.46	-.56	.34	.51	1.00

[a]Lowest N = 181. Correlations of .24 are significant at the < .001 level.

Appendix A

Early Reading Test

Test Data	Tester	Child's Name	ID No.

Picture-Words	Error	Picture Score	Error	Word-Score
STOP				
M & M's				
Kool-aid				
crayons				
baby powder				
	1(51)		3(53)	
EXIT				
coca cola				
Nestle's Quick				
Jello				
telephone				
	2(52)		4(54)	

Spelling

	P A T	T A P E	S A C K	
letters				5(55)
phonemes				5(56)
	words(7)	level(57)		

	T I P	K I T E	S I C K	
letters				8(58)
phonemes				9(59)
	words(10)	level(60)		

Story

stop car	11(61)exact		go cat	exact 14(64)	
stop truck			go dog		
stop bus	12(62)approx		go pig	approx 15(65)	
stop, step, stop			go ge go		
stop for the cat	13 level		go to the car	level 16	

go _____ at _____

ran _____ did _____

but _____ not _____

ten _____ had _____

say _____ may _____

use _____ ate _____

time _____ words _____ letters _____ level ___ time _____ words _____ letters _____ le
17-18(67-68) 19(69) 20(70) 21 22-23(22-73) 24(74) 25(75)

MAKE BELIEVE WORDS

fam _____ lac _____

pag _____ tak _____

caf _____ jav _____

ras _____ gat _____

maz _____ naj _____

san _____ kap _____

time _____ words _____ conson _____ level ___ time _____ words _____ conson _____ lev
27-28(77-78) 29(79) 30(80) 31 32-33(82-83) 34(84) 35(85)

STORY RECALL (STOP)

_____

_____

_____

_____

_____

37 L _____ 87 R _____ 38 C _____ 88 V _____

Look at (last p.) _____

_____

_____

41 C _____

Know about _____

_____

_____

_____

42 C _____ 43 R _____

Think/notice wrong (next-to-last) _____

_____

_____

_____

44 C _____ 45 R _____

STORY RECALL (GO)

_____

_____

_____

_____

_____

39 L _____ 89 R _____ 40 C _____ 90 V ___

Look at (last p.) _____

_____

_____

91 C _____

Know about _____

_____

_____

_____

92 C _____ 93 R _____

Think/notice wrong (next-to-last) _____

_____

_____

_____

94 C _____ 95 R _____

Appendix B

Administering and Scoring the
Chicago

For testing consonant sounds

1. Cover all but the first letter. Ask, WHAT SOUND GOES WITH THIS LETTER? (4 points)
2. Continue in this manner unless child makes an error. If child gives the letter name, say, YES, THAT'S THE LETTER NAME, WHAT IS THE SOUND? (3 points)
3. If child still cannot give the sound, say, THINK OF A WORD THAT STARTS WITH THIS LETTER. (2 points)
4. If child gives word that begins with that letter, then say, NOW TELL ME THE SOUND. (1 point)
5. Continue down the card. Show letters in sequence.

For vowels and endings

1. Cover endings and all but the first vowel.
2. Ask, WHAT SOUND GOES WITH THIS LETTER? (2 points)
3. If child responds correctly, continue. If child gives the name of the letter, say, TELL ME THE ANOTHER SOUND. (1 point)

If the child has failed to identify 2 consonants and one vowel correctly, discontinue testing at this point.

4. After presenting the 4 vowels in order, uncover the first ending, and say, NOW READ THE ENDINGS. (2 points)
5. If child responds correctly, continue. If child read the ending correctly, but had failed to give the correct sound, cover the ending, exposing only the sound, point to the sound and ask, WHAT SOUND GOES WITH THE LETTER? (1 point)

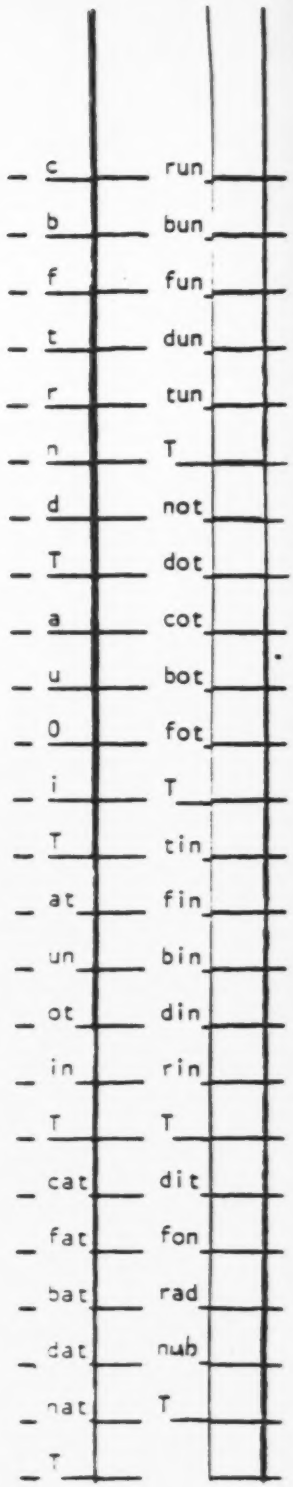
For word families

1. Cover all but the first word. Ask, WHAT WORD? Continue to ask, WHAT WORD? unless child fails to give word (2 points).
   If child fails to give second word after responding correctly to first word, uncover first word again, point to it and say, IF THIS IS CAT, uncover the second word, THEN THIS IS....through the list. (1 point, then for each word identified correctly).
2. If child cannot give first two words in the cat family, give first two words in the run family. If the child fails to identify these two wrods correctly, give the first two words from the not family. If the child fails these two words as well, stop testing.

For nonsense words

1. Ask, WHAT WORD? for each of four words.

4/16/85



Appendix C

Error Detection Test

Absurd Target Word: Practice Item

WE ARE GOING TO PLAY A GAME YOU PROBABLY HAVE NOT PLAYED BEFORE. HERE IS HOW WE PLAY. LISTEN TO ME READ THIS SENTENCE. (Show and read the first card.) SHE WANTS TO PLANT THE CAT. ONE WORD SPOILS THE MEANING. LISTEN AGAIN. SHE WANTS TO PLANT THE CAT. WHICH WORD SPOILS THE MEANING? (If student responds correctly, ask) WHY? VERY GOOD

NOW IT IS YOUR TURN TO READ TO ME.

(Show the first card.)

READ THIS TO ME. Tell child word if identifies incorrectly or hesitates more than 2 seconds.

WHICH WORD SPOILS THE MEANING? WHY?

(Score word-reading errors and target word identification, two scores. If child correctly identifies the target word, then ask) WHY DOES ____ SPOIL THE MEANING?

				Support
1.	We will grow some candy.	Y	N	_____
2.	The cookies will bloom in the spring.	Y	N	_____
3.	The sun needs air and water to grow.	Y	N	_____
4.	Plants need water. Many plants need cakes.	Y	N	_____
5.	Plants grow in yards. Plants grow in pots. Plants grow in toes.	Y	N	_____
6.	Flowers are plants. Trees are plants. Hands are plants.	Y	N	_____
7.	Leaves turn red. Leaves turn yellow. Leaves turn blue.	Y	N	_____
8.	All trees are not the same. Some trees have leaves. Some trees have flowers. But, some trees have fights.	Y	N	_____
9.	All plants have roots. All plants have stems. Some plants have toys.	Y	N	_____
10.	Look into a tiny seed. You will see a root and a stem. You will also see a dog.	Y	N	_____

Impossible Sequence

NOW WE WILL PLAY A NEW GAME. LISTEN TO ME READ. PLANT A BUSH. WATER IT. IT WILL GET SMALLER. WHAT HAPPENED AT THE WRONG TIME? (If student answers incorrectly, re-read the item and then say,) PLANTS DON'T GET SMALLER AFTER YOU PLANT AND WATER THEM, THEY GET BIGGER. LET'S START OVER. I'LL READ AGAIN. PLANT A BUSH. WATER IT. IT WILL GET SMALLER. WHAT HAPPENED AT THE WRONG TIME? (If student answers correctly, then ask,) WHY?

			Support
1. Winter came. The snow fell. Then the flowers grew.	Y	N	_____
2. The lady picked the flower. Then she planted the seed. She watched the seed grow.	Y	N	_____
3. Put a carrot top in a pan of water. First you will see new leaves. Then you will see a root.	Y	N	_____
4. Watch a plant grow. First you see a little plant. Then you see a bigger plant. Then you see only dirt.	Y	N	_____
5. Beans grow like this. First you see young beans on a vine. Then you see seeds. Then you see mature beans.	Y	N	_____
6. Here is how a seed grows into a plant. A seed has a cover on it. It breaks through the seed coat. The tiny plant grows inside the seed.	Y	N	_____

3/27/85

BEST COPY AVAILABLE

Appendix D

3/4/86

Instructions for Administering the Eugene Items

1. WRITE YOUR NAME AT THE TOP OF THESE PAPERS. (Wait.)

Practice item - #0 Read it. Mark it. What did you mark?

2. YOU HAVE AS MUCH TIME AS YOU NEED TO READ THESE STORIES. READ EACH STORY TO YOURSELF. THEN, READ THE QUESTION OR QUESTIONS THAT FOLLOWS THE STORY. CHOOSE THE ANSWER YOU THINK IS BEST AND FILL IN THE OVAL (show on board) RIGHT BEFORE YOUR ANSWER.

3. REMEMBER TO READ AND THINK CAREFULLY. YOU HAVE AS MUCH TIME AS YOU NEED.

0

I have hands.

I have a face.

I tick.

What am I?

- 0 boy
- 0 face
- 0 clock
- 0 clown

1

I have a door.

I have shelves.

I have a light inside me.

I keep things cold.

What am I?

- 0 stove
- 0 shelf
- 0 refrigerator
- 0 remain

2

People wear me.

I help keep them warm.

I have four fingers.

I have a thumb.

I am a --

- 0 hat
- 0 globe
- 0 glove
- 0 scarf

3

I am a vehicle.

I have four wheels.

I have doors.

I have a windshield.

I have seats and a steering wheel.

What am I?

- 0 airplane
- 0 car
- 0 boat
- 0 bicycle

4

I work in an office.

I type.

I write letters.

I answer the phone.

Who am I?

- 0 settle
- 0 nurse
- 0 secretary
- 0 teacher

5

I usually work outside.

I work with wood.

I work with a hammer and nails.

I build things.

Who am I?

- 0 mailman
- 0 carpenter
- 0 carpet
- 0 painter

6

I have branches.

I have leaves.

I am very pretty in the fall.

People like to sit under me when it is hot.

What am I?

- 0 flower
- 0 tent
- 0 tree
- 0 top

7

You use me often.

I am found in the bathroom.

I fit in your hand.

I have a handle.

I have bristles.

What am I?

- 0 hairbrush
- 0 comb
- 0 rag
- 0 soap

8

I have a frame.

I have lenses.

I have earpieces.

What am I?

- O glasses
- O hat
- O window
- O glass

9

One can tell if a samp has deets by running around the samp very quickly. If the samp has deets, the samp will turn a bright blue color. If the samp does not have deets, the samp will remain brown and will not change color. People who live in Ricktara love to eat deets. To collect them, these people have trained their dogs to run around samps. The people watch the samps as the dogs run around them. If a samp turns bright blue, the men rush over and pick the deets from it.

a Samps that have deets sometimes --

- O run around
- O collect
- O like dogs
- O turn bright blue

b Samps turn color if one --

- O hits them
- O runs around them very quickly
- O burns them
- O looks at them

C Deets are loved by the people who --

- O turn blue
- O live in Rickara
- O run around
- O pick berries

d Some people who collect deets use --

O dogs

O samps

O collars

O guns

10

If one goes out on a clear night, he can see the blinkermoble. The blinkermoble is the largest thing that one can see in the night sky. It looks bright yellow or white. Men used to think that the blinkermoble was made of cheese. Men used to think that it was very close to the earth. We know now that it is very far away. The blinkermoble is 240,000 miles from the earth. That is a long, long way. But a few men have made that long trip from the earth to the blinkermoble. They have gone in space ships. Some of these men have gone for a walk on the blinkermoble. They have even sent back television pictures of their walk on the blinkermoble.

a The blinkermoble is --

- O made of cheese
- O close to the earth
- O 2,000 miles from the earth
- O 240,000 miles from the earth

b You can see the blinkermoble --

- O only with glasses
- O at night
- O on a table
- O on top of your television

c The blinkermoble is really —

- O a space ship
- O a long trip
- O the sun
- O the moon

d Who has gone to the blinkermoble?

- O many women
- O a few men
- O the cow that jumped over the moon
- O nobody

11

One dark night, when Sanchas was getting ready to go fishing in his boat, he heard a very strange noise. He turned around and saw a nurzit. It was dark brown and very large -- twice as big as Sanchas. Sanchas did not have a weapon to defend himself. He knew that he could not outrun the nurzit, so he stood there. The nurzit stood up on its hind legs. It sniffed the air. Then it growled. Sanchas could see the huge teeth inside the nurzit's mouth. The nurzit stared at Sanchas for a moment. Then the nurzit went over to a large tree near the shore of the lake. The nurzit climbed the tree. Sanchas could not see what the nurzit was doing in the tree, but Sanchas could hear the sound of bees buzzing around. Sanchas knew that the nurzit was getting honey from a bee hive. Nurzits love honey.

a When Sanchas saw the nurzit, it was --

- O yesterday
- O night
- O morning
- O last year

b Sanchas --

- O had no weapon
- O had a knife
- O could run faster than the nurzit
- O was not afraid

c The nurzit --

- O ate an apple
- O sang a song
- O said hello to Sanchas
- O stood on its hind legs

d When the nurzit was in the tree Sanchas --

- O could not see what the nurzit was doing
- O could not catch fish
- O could not hear bees
- O could not run

e Nurzits love to eat --

- O Sanchas
- O bees
- O honey
- O money

f A nurzit must be a --

- O car
- O cow
- O bear
- O lion

12 Of the four most popular darnacks, the bibble is the fastest. There are two good reasons for its speed. The first is that the bibble has a very large engine. Its engine is almost twice as big as the Rentu's engine. The second reason for the bibble's speed is that it is the lightest darnack. It weighs 500 pounds less than the Rentu. The speed of the bibble was shown in the famous Gant road race last year. All three bibbles crossed the finish line before the first Rentu finished the race.

a Darnacks are --

- O cars
- O light
- O trees
- O gants

b The fastest darnack is the --

- O Rentu
- O Gant
- O road race
- O bibble

C The bibble is --

- O the least popular darnack
- O the slowest darnack
- O the fastest darnack
- O not a darnack

d The bibble has a bigger engine and it is --

O stronger

O lighter

O longer

O prettier

e The bibble weighs 500 pounds less than --

O a truck

O an elephant

O a gant

O a Rentu

13

One should take care of his herns. A baby is not born with herns. The first herns appear when the baby is three or four months old. When the first herns come out the baby is very cranky. He cries and like to bite things. The new herns hurt his mouth. By the time a child is five years old, he has all of his baby herns. Then these herns are pushed outby permanent herns. Some children put their baby herns under their pillow at night. In the morning the hern is gone and a dime is under the pillow. The permanent herns must last one all his life, so he should brush his herns twice a day. He should also go to a dentist twice a year to make sure that he has no cavities in his herns.

a Babies --

- O have herns when they are born
- O don't have herns when they are born
- O have cavities when they are born
- O have permanent herns when they are born

b The first herns appear when the baby is —

- O sitting up
- O biting things
- O three or four months old
- O three or four years old

c When the first herns appear, the baby —

- O is cranky
- O is not cranky
- O has permanent herns
- O is under the pillow

d One should brush his herns --

- O under the pillow
- O twice a year
- O with a hair brush
- O twice a day

e The permanent herns --

- O are baby herns
- O must last one all his life
- O come in twice a day
- O come in twice a year

f One should go to a dentist --

- O twice a day
- O twice a year
- O to get baby herns
- O to get permanent herns

g What do you think herns are?

- O feet
- O eyes
- O teeth
- O diapers

14

Glert is one of the most important things on earth. In some parts of the world, glert is very rare. There are parts of the United States where there is not much glert. These places are called deserts. There are no lakes. People who live in the desert may not see one drop of rain during an entire year. However, most of the earth is covered with glert. The great oceans are filled with glert. Lakes and streams are filled with glert. Even clouds are made up of tiny drops of glert.

a Places where there is not much glert are called --

- O oceans
- O the United States
- O deserts
- O cities

b Lakes and streams are filled with --

- O cats
- O rivers
- O glert
- O boulders

c In the desert it may not rain once in a --

- O day
- O night
- O city
- O year

d What is made of glert?

O clouds

O buildings

O deserts

O rare

e Glert must be --

O sand

O desert

O water

O ocean